

AAI-500 – Team Project Status Update Form

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: 4

Team Leader/Representative: Ali Mohtat

Full Names of Team Members:

1. Ali Mohtat
2. Omar Sagoo

Title of Your Project: The Impact of User Feedback on Drug Success: A Data-Driven Approach to Analyzing Ratings and Conditions

Short Description of Your Project and Objectives:

This project aims to analyze a comprehensive dataset of drug reviews from Drugs.com to evaluate the effectiveness and user perceptions of various medications. By examining user ratings and the medical conditions for which drugs are prescribed, the project seeks to uncover patterns and associations that can inform healthcare professionals and patients about the potential efficacy and usefulness of specific drugs.

Objectives:

1. **Data Exploration:** To explore and preprocess the drug review dataset, identifying key variables such as drug names, conditions, ratings, and useful counts.
2. **Descriptive Analysis:** To provide a summary of user ratings across different drugs and conditions, highlighting the most common associations and trends in the data.
3. **Clustering Analysis:** To implement clustering techniques on the rating and useful count features to group similar drugs based on user feedback and perceived effectiveness.
4. **Visualization:** To visualize the relationships between drugs and conditions through heatmaps and scatter plots, enhancing understanding of user experiences and drug effectiveness.
5. **Insights Generation:** To derive actionable insights from the analysis that can assist healthcare professionals, inform patients, and contribute to further research in pharmacology.

Name of Your Selected Dataset: Drug Reviews (Drugs.com) (from UCI)

Description of Your Selected Dataset (data source, number of variables, size of dataset, etc.):

Data is from UC Irvine machine learning repository

(<https://archive.ics.uci.edu/dataset/462/drug+review+dataset+drugs+com>)

Number of Variables: The dataset contains 7 variables:

Unnamed: 0: An index or identifier for each entry (not utilized for analysis).

drugName: The name of the drug being reviewed (categorical).

condition: The medical condition that the drug is prescribed for (categorical).

review: The textual review provided by users about the drug (text).

rating: The numerical rating given by users, typically on a scale from 1 to 10 (numeric, float).

date: The date of the review submission (object).

usefulCount: The number of users who found the review helpful (numeric, integer).

Size of Dataset: The dataset consists of 161,297 entries, indicating a substantial amount of user feedback across various drugs and conditions. This large sample size provides a robust basis for analysis and helps to identify meaningful patterns and trends in drug effectiveness and user experiences.

Are you using and practicing GitHub as a code hosting platform for version control and collaboration? If yes, provide the link here: <https://github.com/omarsagoo/drug-review-analysis>

How many times have your members met in the last two weeks? 5

List the specific contributions that each team member is providing for the Final Team Project in the table below.

- **NOTE:** ALL students on the team should contribute equally to the Final Team Project.

Team Member 1 (Ali Mohtat)	Team Member 2 (Omar Sagoo)
<ul style="list-style-type: none"> - Data Exploration - Clustering Analysis (50%) - Insights Generation - Final report (50%) - Final presentation (50%) 	<ul style="list-style-type: none"> - git admin - Descriptive Analysis - Clustering Analysis (50%) - Visualization - Final report (50%) - Final presentation (50%)

Comments/ Roadblocks: Nothing to report.